

Robert Nozick

The
Nature
of
Rationality

THE NATURE OF RATIONALITY

Robert Nozick

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

COPYRIGHT © 1993 BY ROBERT NOZICK
PUBLISHED BY PRINCETON UNIVERSITY PRESS, 41 WILLIAM STREET,
PRINCETON, NEW JERSEY 08540
IN THE UNITED KINGDOM: PRINCETON UNIVERSITY PRESS,
CHICHESTER, WEST SUSSEX
ALL RIGHTS RESERVED

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

NOZIK, ROBERT.
THE NATURE OF RATIONALITY / ROBERT NOZICK.
P. CM.
INCLUDES INDEX.
ISBN 0-691-02096-5
ISBN 0-691-00162-6 (PBK.)
1. REASONING. 2. REASON. I. TITLE.
BC177.N69 1993 128'.3—DC20 92-46660

THIS BOOK HAS BEEN COMPOSED IN ADOBE PALATINO

PRINCETON UNIVERSITY PRESS BOOKS ARE PRINTED
ON ACID-FREE PAPER AND MEET THE GUIDELINES FOR
PERMANENCE AND DURABILITY OF THE COMMITTEE ON
PRODUCTION GUIDELINES FOR BOOK LONGEVITY OF
THE COUNCIL ON LIBRARY RESOURCES

THIRD PRINTING, AND FIRST PRINCETON
PAPERBACK PRINTING, 1995

PRINTED IN THE UNITED STATES OF AMERICA

5 7 9 10 8 6 4

**THE NATURE OF
RATIONALITY**

To Carl Hempel
AND TO THE MEMORY OF
Gregory Vlastos

ACKNOWLEDGMENTS

THE FIRST two chapters of this book were originally delivered as Tanner Lectures at Princeton University on November 13 and 15, 1991. I had been a graduate student at Princeton, and the lectures were dedicated, as is this book, to my teachers there. Chapters 1 and 2 are reprinted with the permission of the University of Utah Press from the *Tanner Lectures on Human Values*, vol. 14 (Salt Lake City: University of Utah Press, © 1992). (Some additions and changes have been made in the versions printed here.) First drafts of these two chapters were written at the Rockefeller Foundation Research Center at Bellagio, Italy, in the summer of 1989.

Portions of Chapter 5 constituted the Walter C. Schnackenberg Memorial Lecture, given at Pacific Lutheran University in March 1990. Parts of Chapters 3–5 were given as a Centennial Lecture at the University of Chicago in May 1992.

I am grateful to the discussants of the lectures at Princeton—Gilbert Harman (who also read the complete manuscript), Clifford Geertz, Susan Hurley, and Amos Tversky—and also to Scott Brewer, Eugene Goodheart, David Gordon, Christine Korsgaard, Elijah Millgram, Bill Puka, Tim Scanlon, Howard Sobel, and William Talbott for their very helpful comments and suggestions. Special thanks go to Amartya Sen for many stimulating discussions of this material, inside classes we have taught together and out.

I am very grateful to Laurance Rockefeller for his interest in and generous support of this research project.

I thank my wife, Gjertrud Schnackenberg, who made the years during which this book was written so romantic and loving—and such fun.

INTRODUCTION

THE WORD *philosophy* means the love of wisdom, but what philosophers really love is reasoning. They formulate theories and marshal reasons to support them, they consider objections and try to meet these, they construct arguments against other views. Even philosophers who proclaim the limitations of reason—the Greek skeptics, David Hume, doubters of the objectivity of science—all adduce reasons for their views and present difficulties for opposing ones. Proclamations or aphorisms are not considered philosophy unless they also enshrine and delineate reasoning.

One thing philosophers reason about is reasoning itself. What principles should it obey? What principles must it obey? Aristotle initiated the explicit formulation and study of deductive principles, writers on science and probability theory delineated modes of nondeductive reasoning and support, Descartes attempted to show why we should trust the results of reasoning, Hume questioned the rationality of our doing so, and Kant demarcated what he held to be reason's proper domain. This delineation of reason was not an academic exercise. Discoveries were to be applied: people's reasoning was to be improved, their beliefs and practices and actions made more rational. Inquiring into the rationality of contemporary beliefs and practices carries risks, Socrates discovered. The traditions of a society sometimes do not withstand scrutiny; not everyone wishes to see the implicit examined explicitly. Even the simple consideration of alternatives can seem a corrosive undercutting of what actually exists, an exposure of arbitrariness.

Rationality fixed human distinctiveness, the Greeks held. "Man is a rational animal." The capacity to be rational demarcates humans from other animals and thus defines them. Human specialness has repeatedly been contracted since the Middle Ages—this was the first large statement about intellectual history that I recall reading. Copernicus, Darwin, and Freud taught us that human beings do not occupy a special place in the universe, they are not special in their origin and are not always guided by rational or even consciously known motives. What continued to give humanity some special status, though, is its capacity for rationality. Perhaps we do not consistently exercise this valuable attribute; yet it sets us apart. Rationality provides us with the (potential) power to investigate and discover anything and everything; it enables us to control and direct our behavior through reasons and the utilization of principles.

Rationality therefore is a crucial component of the self-image of the human species, not simply a tool for gaining knowledge or improving our lives and society. Understanding our rationality brings deeper insight into our nature and into whatever special status we possess. The Greeks saw rationality as independent of animality, certainly not its outgrowth. Evolutionary theory makes it possible to see rationality as one among other animal traits, an evolutionary adaptation with a delimited purpose and function.

This perspective can yield important consequences for philosophy, I believe. Rationality has not been merely the philosophers' special love and an important part of their subject matter; it has been their special tool for discovering truth, a potentially unlimited one. (In the *Critique of Pure Reason*, Kant gave reason a humbler function: not to cognize the heart of an independent reality but to know an empirical realm that it partially constitutes and shapes. Still, its valid scope remained extremely large.) If rationality is an evolutionary adaptation with a delimited purpose and function, designed to work in conjunction with other stable facts that it takes for granted and builds upon, but if philosophy is an attempt of unlimited scope to apply reason and to justify rationally every belief and assumption, then we can understand why many of philosophy's traditional problems have turned out to be intractable and resistant to rational resolution. These problems may result from attempts to extend rationality beyond its delimited evolutionary function. I have in mind here the problems of induction, of other minds, of the external world, and of justifying goals. I shall explore the consequences and implications of this evolutionary perspective later on.

In recent years, rationality has been an object of particular criticism. The claim has been put forth that rationality is *biased* because it is a class-based or male or Western or whatever notion. Yet it is part of rationality to be intent on noticing biases, including its own, and controlling and correcting these. (Might the attempt to correct for biases itself be a bias? But if that is a *criticism*, from what quarter does it come? Is there a view that holds that bias is bad but that correcting it is bad too? If it is held to be impossible to eliminate bias, then in what sense does charging bias constitute a criticism? And would such impossibility mean that there is some one particular bias that is intrinsically resistant to elimination or just that not all biases can be eliminated simultaneously?)

Charging a bias in existing standards does not show that one exists. That is done by using reasoning and evidence—hence using our existing standards—to reach the conclusion that these standards them-

selves, in some applications, show some particular specified distortions and biases. It is not sufficient merely to say that we (all) see the world through our conceptual schemes. The question is: In what specific ways, and by what exact mechanisms, do our particular conceptual schemes and standards distort? Once we are shown this, we can begin to make corrections. Of course, our current standards of rationality are not perfect—in what year should we suppose they became so? But they have real virtues, and to show that they are flawed requires rational argumentation of at least the same weight as those standards being attacked. Detecting particular such flaws is the necessary first step toward repairing them and toward formulating the standards of rationality more adequately. So evidence for charges of bias in standards should be welcomed and sought out. Standards of rationality are a means whereby we rise above, or check, our own particular hopes, wishes, and biases. It would be ironic and tragic if the current widespread criticism of standards of rationality had the effect of removing or undercutting one of the major ways through which humanity is able to correct and rise above personal and group bias.

The study of rationality, which is of such great evaluative and practical importance both personally and socially, has gotten transformed into a technical subject. Principles were sharpened to delineate valid reasoning and to capture the patterns of belief and action supported by reasons. Deductive logic was transformed by Gottlob Frege in the late nineteenth century and burst into technical elaboration in the twentieth. Systems of logic were developed and their own properties and limitations were explored using logical techniques. Probability theory led to formal theories of statistical inference, and mathematization permeated attempts to theorize about the rationality of belief and to formulate the rudiments of an inductive logic, or at least of inductive rules of acceptance. A sleek and powerful theory of rational action—decision theory—was developed in this century by mathematicians, economists, statisticians, and philosophers, and now this theory is applied in a wide variety of theoretical and practical contexts. (The apparatus of this theory provides the framework for the formal theory of rational strategic interaction, game theory, the formal theory of social choice and welfare economics, the theory of microeconomic phenomena, and elaborate theories of the political realm.) The relevant literature is sprinkled with, when not wholly engulfed by, forbidding formulas in unfamiliar symbolic notations that are elaborated into mathematical structures. I do not decry this turn. These current theoretical developments are continuous with the earlier motivations and concerns, and they carry the inquiry much further.

This book too will take account of such technicalities and propose some new ones in the two major areas covered by theories of rationality: rationality of decision and rationality of belief. We shall reformulate current decision theory to include the symbolic meaning of actions, propose a new rule of rational decision (that of maximizing decision-value), and then proceed to trace the implications of this rule for the Prisoner's Dilemma and for Newcomb's Problem. The rationality of belief involves two aspects: support by reasons that make the belief credible, and generation by a process that reliably produces true beliefs. (The evolutionary account I offer to explain the puzzling connection between these aspects reverses the direction of Kant's "Copernican Revolution.") I shall propose two rules to govern rational belief: not believing any statement less credible than some incompatible alternative—the intellectual component—but then believing a statement only if the expected utility (or decision-value) of doing so is greater than that of not believing it—the practical component. This twofold structure then is applied to issues about the "ethics of belief" and a new resolution of the "lottery paradox" is proposed. I also shall explore the scope and limits of instrumental rationality, the effective and efficient pursuit of given goals, and propose some new conditions on the rationality of goals. Because rational thinking also encompasses the formulation of new and fruitful philosophical questions and ideas, some heuristics for doing this shall be presented. Thus, this book is awash in technical details needed to push thinking on the fundamental issues of rationality further.

Yet there is some cause for concern. Until recently, questions about rationality had been the common possession of humankind, sometimes discussed in intricate trains of thought—no one could claim that Kant's *Critique of Pure Reason* is an easy book—but, nevertheless, largely accessible to intelligent people willing to make the effort. New thoughts on these questions were part of the general culture; they shaped the terms of discussion and debate, and sometimes even of sensibility (recall how greatly Kant's thought influenced Coleridge). Now things are different—and not just with the topic of rationality.

The most fruitful and interesting lines of inquiry about many topics of fundamental human concern have taken an increasingly technical turn. It is impossible now to discuss these topics adequately without a grasp of these technical developments, of the new questions they open, and of the ways some traditional positions are undercut. When the *Encyclopedia Britannica* recently published its (second) edition of "Great Books of the Western World," this occasioned some public controversy over the representation—or relative lack—of women and minorities, and over the putative elitism of any canon of

great works.* What received no comment, however, was that many of the greatest intellectual works of the twentieth century were omitted, presumably because they were too technical for the intelligent generally educated reader.

The point is not just that interesting thoughts and results have occurred in this century that are inaccessible to large portions of even a well-educated population—that has been true since Newton. Rather, now these ideas concern topics we want and need to understand, topics we think everyone should understand. Yet without some technical familiarity, these topics cannot be understood or intelligently discussed. The very terms of evaluation have become technical.

Let me give some examples of topics that have undergone technical development. (1) The notion of the general welfare (and Rousseau's notion of the "general will") and an understanding of the purposes of democratic voting procedures have been transformed by Kenneth Arrow's Impossibility Theorem. This shows that several extremely natural and desirable conditions, which apparently should be satisfied by any procedure for determining the general welfare or the democratically most preferred alternative, cannot all be satisfied together. Something has to give. (2) Amartya Sen's work on the Paretian liberal paradox shows that a very natural interpretation of the scope of individual rights and liberties, and of how the choices of society should be rationally organized, cannot be easily fit together. These notions need a new structuring. (3) The fundamental nature of the physical world—the structure of space and time—cannot be understood apart from the technicalities (and mathematics) of space-time as presented in general relativity theory. (4) Similarly for the nature of causality and of the independent character of the physical world as these are depicted in the most precise and successful scientific theory we now possess, quantum field theory. (5) Discussion of the nature and status of mathematical truth—since the Greeks, the exemplar of our best and most certain knowledge—has been drastically transformed by Kurt Gödel's incompleteness theorems. (6) The nature of infinity and its various levels is now elaborated and explored in contemporary set theory. (7) Without the theory of how a price mechanism and associated institutions of private property make rational economic calculation possible, and the decades-long theoretical discussion of whether rational calculation was at all possible in a socialist society, one cannot under-

* I myself do not find a uniform edition of the works of many different authors, with the series title emblazoned more prominently than the titles of the individual works or the authors' names, a fitting presentation of the written accomplishments of the mind. It might be useful, however, for a group to publish a *list* of such books and to reprint those not easily available; different groups might publish different lists.

stand why it is that communist societies were so economically ineffective. (8) Concerning aspects of individual rationality and rational interactions among persons, there have been many theoretical advances: decision theory, game theory, probability theory, and theories of statistical inference.

For each of these topics, this century has seen dramatic new results and theories, ones that are difficult to understand or to discuss responsibly without an understanding of the technical structures and details. This is, I realize, a philosopher's list; social and natural scientists would add further topics. That reinforces my point. The common culture of intelligent, educated, and serious people has lost its grip on many topics that are central to understanding and thinking about society or people or the universe at large. The claim that there are complicated scientific factual issues for whose resolution we must turn to experts, experts who perhaps will disagree—for instance, issues about the environmental effects of various practices—is familiar. What is new is this: many of the very terms and concepts of evaluation and understanding that *we* wish to use have themselves become technical.

I raise this issue without a solution to propose. Of course, expositions of these materials are needed for the general reader. But the clearest of these, if it is indeed to convey the essential ideas accurately, will involve some technical descriptions and developments—and thus be limited in its readers. The task is even more difficult for a work that presents and explores new ideas. I do not *want* the topic of rationality to be taken away from the general reader. Yet some ideas can be stated, specified, or defended only in a somewhat technical manner. I have tried to minimize these technical details, or at least confine them to specific sections. For the intellectual health of our society—not to mention the social health of our intellectuals—the fundamental ideas must stay public.

CONTENTS

ACKNOWLEDGMENTS ix

INTRODUCTION xi

I. How to Do Things with Principles 3

Intellectual Functions 3
Interpersonal Functions 9
Personal Functions 12
Overcoming Temptation 14
Sunk Costs 21
Symbolic Utility 26
Teleological Devices 35

II. Decision-Value 41

Newcomb's Problem 41
Prisoner's Dilemma 50
Finer Distinctions: Consequences and Goals 59

III. Rational Belief 64

Cognitive Goals 67
Responsiveness to Reasons 71
Rules of Rationality 75
Belief 93
Bias 100

IV. Evolutionary Reasons 107

Reasons and Facts 107
Fitness and Function 114
Rationality's Function 119

V. Instrumental Rationality and Its Limits 133

Is Instrumental Rationality Enough? 133
Rational Preferences 139
Testability, Interpretation, and Conditionalization 151
Philosophical Heuristics 163
Rationality's Imagination 172

viii CONTENTS

NOTES 183

SUBJECT INDEX 219

INDEX OF NAMES 224

**THE NATURE OF
RATIONALITY**

I

HOW TO DO THINGS WITH PRINCIPLES

WHAT are principles *for*? Why do we hold principles, why do we put them forth, why do we adhere to them? We could instead simply act on whim or the passion of the moment, or we could maximize our own self-interest and recommend that others do the same. Are principles then a constraint upon whim and self-interest, or is adherence to principles a way of advancing self-interest? What functions do principles serve?

Principles of action group actions, placing them under general rubrics; linked actions are then to be viewed or treated in the same way. This generality can serve different functions: intellectual, interpersonal, intrapersonal, and personal. I start with the intellectual.

Intellectual Functions

Consider judicial decisionmaking. In one imaginable system, a judge simply decides a case so as to yield what she thinks is the best or preferable result in that particular case. Another system of judicial decision involves principled decision: a common law judge is to formulate a principle to fit (most or almost all) precedents and a range of hypothetical cases, and then use this principle to decide the current case.* The attempt to formulate an acceptable general principle is a *test* of your judgment about the particular case: is there *some* adequate general principle—a principle that gives the right result in all established cases and obvious hypothetical ones—that also yields the result you want in this case? If you cannot find such a principle, reconsider what result you do want in this case.

Such a procedure is a test of a particular judgment on the assumption that any correct judgment is yielded by *some* true acceptable gen-

* My aim here is to highlight some general features that principles have outside of the legal realm by analogy to some aspects of judicial decision, not to present a complete picture of the functioning of legal institutions. What is illuminating is the analogy between how a current judicial decision is to be yielded by a principle that fits past precedents and how (outside the law) a principle is to yield correct judgments. That within the legal system *stare decisis* is itself a (higher-order) principle of the law that may sometimes conflict or compete with other principles need not concern us now.